

DIVERGENCE OF COMBINATORIAL AVERAGES AND THE UNBOUNDEDNESS OF THE TRILINEAR HILBERT TRANSFORM

CIPRIAN DEMETER

ABSTRACT. We consider multilinear averages in ergodic theory and harmonic analysis and prove their divergence in some range of L^p spaces. This contrasts with the positive behavior exhibited by these averages in a different range, as proved in [5]. We also prove that the trilinear Hilbert transform is unbounded in a similar range of L^p spaces. The underlying principle behind these constructions is stated, setting up the stage for more general results.

1. INTRODUCTION

Multilinear ergodic averages have proved to be a powerful tool in settling problems from combinatorial number theory. This strategy was initiated by Furstenberg in [6]. He gave an ergodic theoretical proof of a theorem of Szemerédi about the existence of arbitrarily long arithmetic progressions in subsets of integers with positive upper density.

Let $\mathbf{X} = (X, \Sigma, m, T)$ be a dynamical system, i.e. a complete probability space (X, Σ, m) endowed with an invertible transformation $T : X \rightarrow X$ such that $mT^{-1} = m$. Furstenberg's method consisted of first proving that

$$\liminf_{N-M \rightarrow \infty} \frac{1}{N-M} \sum_{l=M+1}^N \int f_1(T^l x) f_2(T^{2l} x) \cdots f_n(T^{nl} x) dm(x) > 0,$$

for each positive nonzero functions $f_i \in L^\infty(\mathbf{X})$, and then using a corresponding principle to transfer information to the integers. Subsequently, other combinatorial averages have been investigated by various authors.

Let now $n > 1$, $m \geq 1$ and consider the $(n-1) \times m$ matrix $A = (a_{i,j})_{i=1, j=1}^{n-1, m}$ with integer entries. We will consider the averages

$$\frac{1}{N^m} \sum_{1 \leq n_1, \dots, n_m \leq N} \prod_{i=1}^{n-1} f_i(T^{\sum_{j=1}^m a_{i,j} n_j} x), \quad (1)$$

where f_1, \dots, f_{n-1} are measurable functions on X . When $m = 1$ and $a_{i,1} = i$, we obtain the aforementioned Furstenberg's averages

$$\frac{1}{N} \sum_{1 \leq l \leq N} \prod_{i=1}^{n-1} f_i(T^{il} x). \quad (2)$$

A related object of interest is given by the averages on m dimensional cubes. Some version of them played a key role in Gowers' proof of Szemerédi's theorem [7]. They correspond to the case where $n = 2^m$ and A is a matrix having on each column a vector

AMS subject classification: 37A45, 42B25.

from $V_m = \{0, 1\}^m \setminus \mathbf{0}$. More precisely, if $\epsilon_1, \dots, \epsilon_{2^m-1}$ is the enumeration of V_m in lexicographically increasing order, then the averages on m dimensional cubes are

$$\tilde{K}_m(f_{\epsilon_1}, \dots, f_{\epsilon_{2^m-1}}, N)(x) = \frac{1}{N^m} \sum_{\vec{i} \in \{1, 2, \dots, N\}^m} \prod_{\epsilon \in V_m} f_\epsilon(T^{\vec{i} \cdot \epsilon} x), \quad (3)$$

where $\vec{i} \cdot \epsilon$ denotes the usual dot product.

Both the averages in (2) and those in (3) have been shown to converge in the L^2 -norm when all the functions f_ϵ are in $L^\infty(\mathbf{X})$, a result due to Host and Kra [9] (a different proof for the averages in (2) is due to Ziegler [14]).

As for the almost everywhere convergence of (1), there are two interesting lines of investigation. On the one hand, it is legitimate to ask whether convergence holds for L^∞ functions. Bourgain proved in [3] using Fourier analysis on the torus that the averages in (2) converge almost everywhere for L^∞ functions, in the case $n = 3$. For larger n only partial results are known, for transformations T with nice spectral properties [2], [13]. Assani [1] used the semi-norms introduced by Host and Kra in [9] to prove the almost everywhere convergence of the averages on cubes for arbitrary m , again for bounded functions.

Once pointwise convergence is established for a dense class of functions (like $L^\infty(\mathbf{X})$), the extension of this convergence to other L^p spaces becomes synonymous with the existence of a maximal inequality for the associated maximal operator. In [5], a very general result is proved, showing the boundedness of the maximal operator

$$T_{A, \mathbf{X}}^*(f_1, \dots, f_{n-1})(x) = \sup_{N \geq 1} \frac{1}{N^m} \sum_{|n_1|, \dots, |n_m| \leq N} \prod_{i=1}^{n-1} |f_i(T^{\sum_{j=1}^m a_{i,j} n_j} x)|$$

in some range of L^p spaces depending on the rank properties of the extended matrix $\mathbb{E}(A)$ defined as

$$\mathbb{E}(A) = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} & 1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} & 1 \\ \dots & \dots & \dots & \dots & \dots \\ a_{n-1,1} & a_{n-1,2} & \dots & a_{n-1,m} & 1 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

Our main objective here is to contrast this result with some negative ones, showing that divergence may occur for some functions in L^p , with p sufficiently close to 1, due to the maximal operator above failing to be bounded. One of the important aspects of these counterexamples is that they give an explanation to why the time frequency methods used in [5] can not be used all the way up to L^1 .

One can also look at the averages in (1) from a different angle and analyze their real variable analog

$$\frac{1}{\varepsilon^m} \int_{|t_1|, \dots, |t_m| \leq \varepsilon} \prod_{i=1}^{n-1} f_i(x + \sum_{j=1}^m a_{i,j} t_j) d\vec{t}, \quad (4)$$

where f_1, \dots, f_{n-1} are arbitrary measurable functions on \mathbb{R} . The question of norm convergence for bounded functions is trivial in this context (one uses approximation and the immediate result for the span of the characteristic functions of intervals). However, the

study of the pointwise convergence of the averages in (1) is intimately connected to that of the averages in (4). Indeed, a maximal inequality (or the failure of it) in one setting is equivalent to the existence (or failure) of a maximal inequality for the corresponding maximal operator in the other context. The transition back and forth between the reals and a dynamical system is mediated by the integers, which can be regarded as a copy of each individual orbit sitting naturally inside the reals. We will do all the constructions in the real variable setting and then transfer them to the averages in (1).

More interestingly, our counterexamples also apply with no modification to the case of the corresponding singular integral operators. In particular, we show that the trilinear Hilbert transform is unbounded in some range of exponents.

Most of the results from the following section have been announced in [5].

2. MAIN RESULTS

We start with the divergence of Furstenberg's nonconventional averages (2) for functions in L^p spaces, with p close to L^1 . We will then sketch the more delicate argument for the averages on cubes.

Theorem 2.1. *Define $p_0 = 1 + \frac{\log_6 2}{1 + \log_6 2}$ and consider $p < p_0$. In every ergodic dynamical system $\mathbf{X} = (X, \Sigma, \mu, T)$ there are three functions $F, G, H \in L^p(\mathbf{X})$ such that*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N F(T^n x)G(T^{2n} x)H(T^{3n} x) = \infty$$

for μ a.e. $x \in X$.

Proof The idea of the proof is to construct “small” subsets of integers A_0 and B_0 , such that $2B_0 - A_0$ is also “small” while $2A_0 - B_0$ is “large”. This choice is dictated by the geometry of the averages under investigation, as explained below.

For a fixed $k \geq 1$ consider the following subsets of $[-1, 1]$:

$$\begin{aligned} A &= \left\{ \sum_{i=1}^k \frac{a_i}{12^i} + z : a_i \in A_0 := \{-4, -2, 0\}, 0 \leq z \leq \frac{1}{2 \times 12^k} \right\} \\ B &= \left\{ \sum_{i=1}^k \frac{b_i}{12^i} + z : b_i \in B_0 := \{0, 1, 2, 3\}, 0 \leq z \leq \frac{1}{2 \times 12^k} \right\} \\ C &= \left\{ \sum_{i=1}^k \frac{2b_i - a_i}{12^i} + z : a_i \in \{-4, -2, 0\}, b_i \in \{0, 1, 2, 3\}, 0 \leq z \leq \frac{1}{2 \times 12^k} \right\} \\ &= \left\{ \sum_{i=1}^k \frac{c_i}{12^i} + z : c_i \in \{0, 2, 4, 6, 8, 10\}, 0 \leq z \leq \frac{1}{2 \times 12^k} \right\} \\ D &= \left\{ \sum_{i=1}^k \frac{2a_i - b_i}{12^i} + z : a_i \in \{-4, -2, 0\}, b_i \in \{0, 1, 2, 3\}, 0 \leq z \leq \frac{1}{8 \times 12^k} \right\} \\ &= \left\{ \sum_{i=1}^k \frac{d_i}{12^i} + z : d_i \in \{-11, -10, \dots, -1, 0\}, 0 \leq z \leq \frac{1}{8 \times 12^k} \right\}. \end{aligned}$$

We claim that

$$m \left\{ x \in [-1, 0] : \int_0^1 1_A(x+t)1_B(x+2t)1_C(x+3t)dt \geq \frac{1}{8 \times 12^k} \right\} \geq \frac{1}{8}, \quad (5)$$

where m denotes the Lebesgue measure. Indeed, for each $x \in D$ of the form $x = \sum_{i=1}^k \frac{2a_i - b_i}{12^i} + z$ (for some i), with $0 \leq z \leq \frac{1}{8 \times 12^k}$, and for each t of the form $t = \sum_{i=1}^k \frac{b_i - a_i}{12^i} + z'$ (for the same i), with $0 \leq z' \leq \frac{1}{8 \times 12^k}$, we immediately see that $x+t \in A$, $x+2t \in B$ and $x+3t \in C$. It now suffices to note that $m(D) = \frac{1}{8}$, and the claim follows. From here we deduce, by the definition of Riemann integral, that for some sufficiently large N_k (independent of x) we have

$$m \left\{ x \in [-1, 0] : \frac{1}{N_k} \sum_{n=1}^{N_k} 1_A(x + \frac{n}{N_k})1_B(x + \frac{2n}{N_k})1_C(x + \frac{3n}{N_k}) \geq \frac{1}{16 \times 12^k} \right\} \geq \frac{1}{8}. \quad (6)$$

Now consider a $p < p_0$ and the dynamical system $\mathbf{X}_k = ([-1, 1], B, m_1, T_k)$, where B is the restriction of the Lebesgue algebra, m_1 is the normalized Lebesgue measure (i.e. $m_1([-1, 1]) = 1$) and $T_k(x) = x + \frac{1}{N_k}$, with the addition considered modulo the interval $[-1, 1]$. We can now rephrase (6) as

$$m_1 \left\{ x \in [-1, 1] : \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N 1_A(T_k^n x)1_B(T_k^{2n} x)1_C(T_k^{3n} x) \geq \frac{1}{16 \times 12^k} \right\} \geq \frac{1}{16}. \quad (7)$$

A more useful way of stating this, upon noting that $\|1_A\|_{L^p(\mathbf{X}_k)} = \frac{1}{(4 \times 4^k)^{1/p}}$, $\|1_B\|_{L^p(\mathbf{X}_k)} = \frac{1}{(4 \times 3^k)^{1/p}}$ and $\|1_C\|_{L^p(\mathbf{X}_k)} = \frac{1}{(4 \times 2^k)^{1/p}}$, is

$$\sup_{\substack{\|f\|_{L^p(\mathbf{X}_k)} = \frac{1}{(4 \times 4^k)^{1/p}} \\ \|g\|_{L^p(\mathbf{X}_k)} = \frac{1}{(4 \times 3^k)^{1/p}} \\ \|h\|_{L^p(\mathbf{X}_k)} = \frac{1}{(4 \times 2^k)^{1/p}}}} m_1 \left\{ x \in [-1, 1] : \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f(T_k^n x)g(T_k^{2n} x)h(T_k^{3n} x) \geq \frac{1}{16 \times 12^k} \right\} \geq \frac{1}{16}.$$

By using the transference principle proved in the Appendix, it follows that there exist $f_k \in L^p(\mathbf{X})$, $g_k \in L^p(\mathbf{X})$ and $h_k \in L^p(\mathbf{X})$ with $\|f_k\|_{L^p(\mathbf{X})} = \frac{1}{(4 \times 4^k)^{1/p}}$, $\|g_k\|_{L^p(\mathbf{X})} = \frac{1}{(4 \times 3^k)^{1/p}}$, $\|h_k\|_{L^p(\mathbf{X})} = \frac{1}{(4 \times 2^k)^{1/p}}$, such that

$$\mu \left(E_k := \left\{ x \in X : \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f_k(T^n x)g_k(T^{2n} x)h_k(T^{3n} x) \geq \frac{1}{32 \times 12^k} \right\} \right) \geq \frac{1}{16}. \quad (8)$$

Define now

$$\begin{aligned} f &= \sum_{k=1}^{\infty} \frac{(4 \times 4^k)^{1/p}}{k^2} f_k \\ g &= \sum_{k=1}^{\infty} \frac{(4 \times 3^k)^{1/p}}{k^2} g_k \\ h &= \sum_{k=1}^{\infty} \frac{(4 \times 2^k)^{1/p}}{k^2} h_k. \end{aligned}$$

Note that $f, g, h \in L^p(\mathbf{X})$. Moreover, there exists a set $X_0 \subset X$ of positive measure such that each x from X_0 belongs to infinitely many E_k 's. For each such x and k we have

$$\sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f(T^n x)g(T^{2n} x)h(T^{3n} x) \geq \frac{(4 \times 4^k)^{1/p}(4 \times 3^k)^{1/p}(4 \times 2^k)^{1/p}}{32k^6 \times 12^k}.$$

Due to our choice of p , the sequence on the left above goes to ∞ . Since

$$\bigcup_{j=1}^{\infty} T^j(X_0) = X$$

up to sets of μ measure 0, it follows that

$$F := \sum_{j=1}^{\infty} \frac{1}{j^2} T^j f$$

$$G := \sum_{j=1}^{\infty} \frac{1}{j^2} T^j g$$

$$H := \sum_{j=1}^{\infty} \frac{1}{j^2} T^j h$$

satisfy the requirement of Theorem (2.1). ■

Remark 2.2. No negative result can be proved with these techniques for the bilinear averages, the reason being that $x + t$ and $x + 2t$ are linearly independent monomials in $\mathbb{R}[x, t]$. The result of Lacey [12] shows that they do behave well for some range of indices, however opposite behavior is quite anticipated near L^1 .

Remark 2.3. The index p_0 in the above theorem can be pushed as close as desired to $\frac{3}{2}$, if one considers instead the more general trilinear ergodic averages

$$\frac{1}{N} \sum_{n=1}^N F(T^{a_1 n} x)G(T^{a_2 n} x)H(T^{a_3 n} x),$$

with $a_1 \neq a_2 \neq a_3 \in \mathbf{Z}$. (the choice for a_i will depend on how close p_0 is to $\frac{3}{2}$) This follows as a consequence of the results obtained by Christ in [4], which were also the main inspiration for our investigation here. On the other hand, it is a consequence of Lacey's result [12] and multilinear interpolation, as explained in [5], that p_0 can be at most 2.

Remark 2.4. The result in Theorem 2.1 immediately proves the divergence of Furstenberg's averages for any $n \geq 3$ (just choose the remaining functions to be identically equal to one).

Interestingly, the constructions above also prove the unboundedness of the trilinear Hilbert transform. This operator is initially defined for piecewise continuous functions f_i with finite support as follows:

$$H_3(f_1, f_2, f_3)(x) = p.v. \int f_1(x+t)f_2(x+2t)f_3(x+3t) \frac{dt}{t}$$

The bilinear version H_2 of this -also known as the bilinear Hilbert transform

$$H_2(f_1, f_2)(x) = p.v. \int f_1(x+t) f_2(x+2t) \frac{dt}{t},$$

was proved to be bounded by Lacey and Thiele [10], [11] in some range of exponents. The type of time-frequency analysis involved in their argument seems insufficient at the moment to address any positive bounds for H_3 .

Theorem 2.5. *Define $p_0 = 1 + \frac{\log_6 2}{1+\log_6 2}$ and consider $p < p_0$. Then the inequality*

$$\|H_3(f_1, f_2, f_3)\|_{p/3} \leq C \|f_1\|_p \|f_2\|_p \|f_3\|_p \quad (9)$$

fails to hold with a universal constant C , independent of f_i .

Proof It suffices to note that with the notation in the proof of Theorem 2.1 we have the following consequence of (5)

$$m \left\{ x \in [-1, 0] : \int_{-\infty}^{\infty} 1_A(x+t) 1_B(x+2t) 1_C(x+3t) \frac{dt}{t} \geq \frac{1}{8 \times 12^k} \right\} \geq \frac{1}{8}.$$

To see this, observe that since x is negative and since B consists only of positive numbers we have

$$H_3(1_A, 1_B, 1_C)(x) = \int_0^{\infty} 1_A(x+t) 1_B(x+2t) 1_C(x+3t) \frac{dt}{t} \geq \int_0^1 1_A(x+t) 1_B(x+2t) 1_C(x+3t) dt$$

for each $x \in [-1, 0]$. Note also that as before $\|1_A\|_{L^p} = \frac{1}{(4 \times 4^k)^{1/p}}$, $\|1_B\|_{L^p} = \frac{1}{(4 \times 3^k)^{1/p}}$, $\|1_C\|_{L^p} = \frac{1}{(4 \times 2^k)^{1/p}}$, and $\|H_3(1_A, 1_B, 1_C)\|_{p/3} \geq \frac{1}{8^{3/p}} \frac{1}{(8 \times 12^k)}$. By using $f_1 = 1_A$, $f_2 = 1_B$, $f_3 = 1_C$, a simple computation now shows that the constant C in (9) will go to ∞ as k goes to ∞ . ■

The following analog of Theorem 2.1 holds for the averages on cubes (3).

Theorem 2.6. *For each $m \geq 3$, for each $p < \frac{2^{m-1}+1}{m+1}$, given any ergodic dynamical system $\mathbf{X} = (X, \Sigma, \mu, T)$, there are $2^m - 1$ functions $F_\epsilon \in L^p(\mathbf{X})$, $\epsilon \in V_m$, satisfying*

$$\limsup_{N \rightarrow \infty} \tilde{K}_m(F_{\epsilon_1}, \dots, F_{\epsilon_{2^m-1}}, N)(x) = \infty$$

for μ a.e. $x \in X$.

Proof As before, we first analyze the analog of the above averages on \mathbb{R} , defined by

$$K_m(f_{\epsilon_1}, \dots, f_{\epsilon_{2^m-1}}, \lambda)(x) := \frac{1}{\lambda^m} \int_0^\lambda \dots \int_0^\lambda \prod_{\epsilon \in V_m} F_\epsilon(x + \vec{t} \cdot \epsilon) dt_1 \dots dt_m,$$

where $\vec{t} = (t_1, \dots, t_m)$. Fix a $k \geq 1$. Define first

$$\tilde{A}_{(1, \dots, 1_{j-1}, 0, 1_{j+1}, \dots, 1)} = \left\{ \sum_{i=1}^k \frac{a_i}{(2^{m+1})^i} : a_i \in \{0, 2^{j-1}\} \right\}$$

for each $1 \leq j \leq m$, (which will be denoted with B_j , for simplicity in future reference), and also

$$\tilde{A}_{(1,\dots,1)} = \left\{ \sum_{i=1}^k \frac{a_i}{(2^{m+1})^i} : a_i \in \{0, -\frac{2^m}{m-1}\} \right\},$$

(which will be denoted with B , for simplicity in future reference). All sets \tilde{A} are indexed by vectors in V_m . We uniquely determine the remaining \tilde{A}_ϵ 's as being the minimal sets which satisfy the constraint: for each $x, t_1, \dots, t_m \in \mathbb{R}$ satisfying

$$x + t_1 + \dots + t_{j-1} + t_{j+1} + \dots + t_m \in \tilde{A}_{(1,\dots,1_{j-1},0,1_{j+1},\dots,1)} \quad (10)$$

for each $1 \leq j \leq m$, and

$$x + t_1 + \dots + t_m \in \tilde{A}_{(1,\dots,1)}, \quad (11)$$

we also have that

$$x + \vec{t} \cdot \epsilon \in \tilde{A}_\epsilon$$

for all the remaining $\epsilon \in V_m$. A simple inspection shows that the monomials in x and t from (10) and (11) are linearly independent. Also, for each $1 \leq l \leq m-2$ and for each $\epsilon \in V_m$ with exactly l nonzero entries, \tilde{A}_ϵ can be written as a linear combination of exactly $m+1-l$ of the $m+1$ \tilde{A}_ϵ 's from (10) and (11). For example

$$\tilde{A}_{(1,0,\dots,0)} = B_1 + B_3 + \dots + B_m - (m-2)B,$$

$$\tilde{A}_{(1,1,0,\dots,0)} = B_1 + B_4 + \dots + B_m - (m-3)B, \text{ etc.}$$

Since each B_j and B have 2^k elements, we conclude that for each $1 \leq l \leq m-2$, $\binom{m}{l}$ of the \tilde{A}_ϵ 's will have at most $2^{(m-l+1)k}$ elements. For $l=m-1$, we have m such \tilde{A}_ϵ (i.e. the B_j 's) with 2^k elements each, while for $l=m$ there exists one such set (namely B) with 2^k elements. Note also that (10) and (11) imply that

$$x \in \tilde{C} := B_1 + \dots + B_m - (m-1)B \quad (12)$$

$$t_j \in \tilde{C}_j := B - B_{j+1} \quad (13)$$

for each $1 \leq j \leq m$, with the obvious agreement that $B_{m+1} := B_1$. Also, for each $x = b_1 + \dots + b_m - b \in \tilde{C}$ there exists a unique m tuple $\vec{t} = (t_1, \dots, t_m) \in \prod_{j=1}^m \tilde{C}_j$, namely

$$(t_1, \dots, t_m) = (b - b_2, b - b_3, \dots, b - b_1), \quad (14)$$

such that (10) and (11) hold.

One can easily see that

$$\tilde{C} = \left\{ \sum_{i=1}^k \frac{c_i}{(2^{m+1})^i} : c_i \in \{0, 1, \dots, 2^{m+1}-1\} \right\}$$

has $2^{k(m+1)}$ elements, and each two of them are separated by at least $\frac{1}{2^{k(m+1)}}$. Define now the sets

$$A_\epsilon = \tilde{A}_\epsilon + [0, \frac{1}{2^{k(m+1)}}]$$

for each $\epsilon \in V_n$ and

$$C = \tilde{C} + [0, \frac{1}{(m+1)2^{k(m+1)}}]$$

Take now some $x = b_1 + \dots + b_m - b + z \in C$, with $0 \leq z \leq \frac{1}{(m+1)2^{k(m+1)}}$. The above discussion shows that for each $t_j \in b - b_{j+1} + [0, \frac{1}{(m+1)2^{k(m+1)}}]$ we have that for each $\epsilon \in V_m$

$$x + \vec{t} \cdot \epsilon \in A_\epsilon.$$

Define the functions

$$f_{k,\epsilon} = 1_{A_\epsilon}$$

and note that

$$\|f_{k,\epsilon}\|_{L^p(\mathbb{R})} \leq \left(\frac{|\tilde{A}_\epsilon|}{2^{k(m+1)}} \right)^{\frac{1}{p}},$$

and also that

$$m \left\{ x : \int_{-1}^0 \dots \int_{-1}^0 \prod_{\epsilon \in V_m} f_{k,\epsilon}(x + \vec{t} \cdot \epsilon) d\vec{t} \geq \frac{1}{[(m+1)2^{k(m+1)}]^n} \right\} \geq \frac{2^{k(m+1)}}{(m+1)2^{k(m+1)}} = \frac{1}{m+1}.$$

The argument continues like in Theorem 2.1, and hence we are guaranteed the negative result as long as the sequence

$$\frac{1}{[(m+1)2^{k(m+1)}]^m} \prod_{\epsilon \in V_m} \left(\frac{2^{k(m+1)}}{|\tilde{A}_\epsilon|} \right)^{\frac{1}{p}}$$

diverges to ∞ . Using the discussion on the sizes of the \tilde{A}_ϵ 's from before, it suffices to have

$$\frac{1}{[(m+1)2^{k(m+1)}]^m} \left(\prod_{l=1}^{m-2} 2^{lk \binom{m}{l}} 2^{km(m+1)} \right)^{\frac{1}{p}} \rightarrow \infty.$$

This is easily seen to happen for $p < 1 + \frac{\sum_{l=1}^{m-2} l \binom{m}{l}}{m(m+1)} = \frac{2^{m-1} + 1}{m+1}$. \blacksquare

Remark 2.7. Note that a negative result is produced as soon as one realizes that at least one of those \tilde{A}_ϵ 's for which ϵ has at most $m-2$ nonzero entries, has cardinality considerably less than $2^{k(m+1)}$.

As in the case of the bilinear averages analyzed earlier, this type of constructions can not prove divergence for the averages on squares ($m = 2$), but do offer valuable information regarding their degenerate analog

$$\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N f(T^{2i}x)g(T^{2j}x)h(T^{i+j}x). \quad (15)$$

We note first that as a result of multilinear interpolation, the maximal operator associated with these averages maps boundedly $L^{p_1} \times L^{p_2} \times L^{p_3}$ into $L^{p'_4}$, where $\frac{1}{p'_4} = \frac{1}{p_1} + \frac{1}{p_2} + \frac{1}{p_3}$, whenever $p'_4 > \frac{1}{2}$ and $1 < p_1, p_2, p_3$. Indeed, the boundedness is immediate in the case $p_1 = p_2 = 1 + \epsilon$, $p_3 = \infty$, while the other five permutations can be seen similarly, after some change of variables. This is sharp as the following proposition shows.

Proposition 2.8. *The maximal operator associated with the averages (15) fails to map boundedly $L^{p_1} \times L^{p_2} \times L^{p_3}$ into $L^{p'_4}$, whenever $p'_4 < \frac{1}{2}$.*

Proof To see this, it suffices to repeat the argument in Theorem 2.1 for $f = g = h = 1_A$ where $A := \{z \in \mathbb{R} : |z| \leq \frac{1}{M}\}$, for large M . The maximal inequality will have a constant that will go to ∞ as M goes to ∞ . The details are left to the reader. ■

The above proposition combined with the positive result in [5] for the nondegenerate squares in the range $p'_4 > \frac{2}{5}$, makes the important point that the almost everywhere behavior on various L^p spaces ($p < \infty$) of the averages

$$\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N f(T_1^i x)g(T_2^j x)h(T_3^{i+j} x)$$

associated with general commuting measure preserving transformations T_1, T_2, T_3 , is very sensitive to the relation between the T_i 's. This is unlike the case where $f, g, h \in L^\infty(\mathbf{X})$, for which the a.e. convergence holds for any (not necessarily commuting) transformations T_1, T_2, T_3 (see [1]).

The main ingredient behind all the negative results proved here is the fact that the monomials $x + \sum_{j=1}^m a_{i,j} t_j, 1 \leq i \leq n-1$ are linearly dependent in $\mathbb{R}[x, t_1, \dots, t_l]$. This allows for the functions involved to be simultaneously large, hence making the maximal operator big on a relevant set of x 's. Similar constructions can be made for various other averages of this type, however it is not clear whether this approach can always be applied. Here is a brief account on what the main difficulty is. Let r be the smallest number for which one can find r linearly dependent monomials as above, and assume for simplicity that these correspond to $i \in \{1, \dots, r\}$. Then we can always find at least $r-2$ monomials among $\sum_{j=1}^m a_{i,j} t_j, 1 \leq i \leq r$ which are linearly independent in $\mathbb{R}[t_1, \dots, t_l]$. One type of situation occurs when only $r-2$ such monomials exist, say the ones with $i \in \{1, \dots, r-2\}$. By performing a suitable change of variables and by ignoring the influence of $f_i, i \geq r+1$, it suffices to show that the maximal operator

$$\sup_{\epsilon > 0} \frac{1}{\epsilon^{r-2}} \int_{|t_1|, \dots, |t_{r-2}| \leq \epsilon} \prod_{i=1}^{r-2} |f_i(x + t_i)| \cdot |f_{r-1}(x + \sum_{i=1}^{r-2} b_i t_i)| |f_r(x + \sum_{i=1}^{r-2} c_i t_i)| dt_1 \dots dt_{r-2}, \quad (16)$$

fails to be bounded in some appropriate range of exponents. The coefficients b_i, c_i are arbitrary integers with the property that the monomial $x + \sum_{i=1}^{r-2} c_i t_i$ can be written a linear combination of $x + t_i, 1 \leq i \leq r-2$ and $x + \sum_{i=1}^{r-2} b_i t_i$. In this scenario, the constructions would have to be sensitive to the arithmetics of the b_i 's and c_i 's. The second type of situation occurs when we can find $r-1$ linearly independent monomials in $\mathbb{R}[t_1, \dots, t_l]$ (the degenerate case). A similar reasoning shows that in this case it suffices to analyze the maximal operator

$$\sup_{\epsilon > 0} \frac{1}{\epsilon^{r-1}} \int_{|t_1|, \dots, |t_{r-1}| \leq \epsilon} \prod_{i=1}^{r-1} |f_i(x + t_i)| \cdot |f_r(x + \sum_{i=1}^{r-1} b_i t_i)| dt_1 \dots dt_{r-1}, \quad (17)$$

where the coefficients b_i are arbitrary integers with the property that the monomial $x + \sum_{i=1}^{r-2} b_i t_i$ can be written a linear combination of $x + t_i, 1 \leq i \leq r-1$. In this case, the

same construction like the one described above for the averages on degenerate squares will produce divergence results for some $f_i \in L^p, 1 \leq i \leq r$, whenever $p < \frac{r}{r-1}$.

3. APPENDIX

For a complete nonatomic probability space (X, Σ, μ) , denote by $C(X)$ the family of all the invertible μ -measure preserving transformations T of X . Equip $C(X)$ with the topology of weak convergence, in which $T_s \rightarrow T$ if and only if $\mu(T_s A \Delta T A) \rightarrow 0$ for each $A \in \Sigma$. If a second complete nonatomic probability space (Y, F, ν) is present, we will denote by $C(Y, X)$ the set of all invertible, bimeasurable transformations $\beta : Y \rightarrow X$ which take measure ν to measure μ . The following result is due to Halmos [8], and is the key to the proof of the transference lemma 3.2:

Theorem 3.1. *Given an ergodic dynamical system $\mathbf{Y} = (Y, F, \nu, S)$ and a complete nonatomic probability space (X, Σ, μ) , the set*

$$\{\beta S \beta^{-1}, \beta \in C(Y, X)\}$$

is dense in the weak topology of $C(X)$.

Lemma 3.2 (Transference Principle). *For a given dynamical system $\mathbf{X}_T = (X, \Sigma, \mu, T)$, $p \geq 1$ and positive constants a, b, c, λ , define*

$$\gamma(\mathbf{X}_T, a, b, c, \lambda) = \sup_{\substack{\|f\|_{L^p(\mathbf{X})}=a \\ \|g\|_{L^p(\mathbf{X})}=b \\ \|h\|_{L^p(\mathbf{X})}=c}} \mu \left\{ x \in X : \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f(T^n x) g(T^{2n} x) h(T^{3n} x) > \lambda \right\}.$$

Then for any ergodic dynamical system $\mathbf{Y}_S = (Y, \mathcal{F}, \nu, S)$ we have

$$\gamma(\mathbf{X}_T, a, b, c, \lambda) \leq \gamma(\mathbf{Y}_S, a, b, c, \lambda).$$

The first goal here is to show that for each $0 \leq r \leq 1$,

$$S_r = \{\sigma \in C(X) : \gamma(\mathbf{X}_\sigma, a, b, c, \lambda) \leq r\}$$

is closed in the weak topology of $C(X)$. Let σ_s be a net in $C(X)$ converging weakly to some σ , such that $\gamma(\mathbf{X}_{\sigma_s}, a, b, c, \lambda) \leq r$. Take $f, g, h \in L^p(\mathbf{X}_T)$ with $\|f\|_{L^p(\mathbf{X}_T)} = a$, $\|g\|_{L^p(\mathbf{X}_T)} = b$ and $\|h\|_{L^p(\mathbf{X}_T)} = c$. From the definition of the weak topology, $f \circ \sigma_s \rightarrow f \circ \sigma$ in the norm of $L^p(\mathbf{X}_T)$ and similarly for g and h . There exists a subnet indexed by \mathbf{Z}_+ , which we will denote by (σ_l) , such that $f \circ \sigma_l \rightarrow f \circ \sigma$, $g \circ \sigma_l \rightarrow g \circ \sigma$ and $h \circ \sigma_l \rightarrow h \circ \sigma$, in the L^p -norm. Choose a subset $X_0 \subset X$ of full measure such that $\lim_{l \rightarrow \infty} f(\sigma_l x) = f(\sigma x)$ and simultaneously for g and h , for each $x \in X_0$. This is easily seen to imply that

$$\sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f(\sigma^n x) g(\sigma^{2n} x) h(\sigma^{3n} x) \leq \liminf_{l \rightarrow \infty} \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f(\sigma_l^n x) g(\sigma_l^{2n} x) h(\sigma_l^{3n} x)$$

for each $x \in X_0$. So

$$\begin{aligned}
& \mu\{x \in X : \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f(\sigma^n x)g(\sigma^{2n} x)h(\sigma^{3n} x) > \lambda\} \\
& \leq \mu\{x \in X : \liminf_{l \rightarrow \infty} \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f(\sigma_l^n x)g(\sigma_l^{2n} x)h(\sigma_l^{3n} x) > \lambda\} \\
& = \lim_{m \rightarrow \infty} \mu\{x \in X : \inf_{l \geq m} \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f(\sigma_l^n x)g(\sigma_l^{2n} x)h(\sigma_l^{3n} x) > \lambda\} \\
& \leq \lim_{m \rightarrow \infty} \mu\{x \in X : \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f(\sigma_m^n x)g(\sigma_m^{2n} x)h(\sigma_m^{3n} x) > \lambda\} \\
& \leq r.
\end{aligned}$$

The fact that S_r is closed follows immediately.

For the last part of the proof, consider an arbitrary system $\mathbf{X}_T = (X, \Sigma, \mu, T)$. Take arbitrary functions $f, g, h \in L^p(\mathbf{X}_T)$ with L^p norms equal to a, b and c , respectively. Theorem 3.1 guarantees the existence of a net of transformations $(\beta_s) \subset C(Y, X)$ such that $\beta_s S \beta_s^{-1} \rightarrow T$ in the weak topology. It is an easy verification that

$$\begin{aligned}
& \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f \circ \beta_s(S^n \beta_s^{-1} x)g \circ \beta_s(S^{2n} \beta_s^{-1} x)h \circ \beta_s(S^{3n} \beta_s^{-1} x) \\
& = \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f((\beta_s S \beta_s^{-1})^n x)g((\beta_s S \beta_s^{-1})^{2n} x)h((\beta_s S \beta_s^{-1})^{3n} x)
\end{aligned}$$

for μ a.e. x . The functions $f \circ \beta_s$, $g \circ \beta_s$ and $h \circ \beta_s$ have the same norms a , b and c respectively in $L^p(\mathbf{Y}_S)$, hence

$$\mu\{x \in X : \sup_{N \geq 1} \frac{1}{N} \sum_{n=1}^N f((\beta_s S \beta_s^{-1})^n x)g((\beta_s S \beta_s^{-1})^{2n} x)h((\beta_s S \beta_s^{-1})^{3n} x) > \lambda\} \leq \gamma(\mathbf{Y}_S, a, b, c, \lambda).$$

Since f, g, h were arbitrary, we get that for each s ,

$$\gamma(\mathbf{Y}_{\beta_s S \beta_s^{-1}}, a, b, c, \lambda) \leq \gamma(\mathbf{Y}_S, a, b, c, \lambda)$$

The fact that $S_{\gamma(\mathbf{Y}_S, a, b, c, \lambda)}$ is closed in $C(X)$ finishes the proof.

Remark 3.3. The obvious analog of the above Transference Principle also holds, with no essential modifications in the proof, for each of the averages in (1).

REFERENCES

- [1] I. Assani. Pointwise convergence of ergodic averages along cubes, preprint.
- [2] I. Assani. Multiple recurrence and almost sure convergence for weakly mixing dynamical systems. *Israel J. Math.* **103** (1998), 111-124.
- [3] J. Bourgain. Double recurrence and almost sure convergence. *J. Reine Angew. Math.* **404** (1990), 140-161.
- [4] M. Christ. On certain elementary trilinear operators. *Math. Res. Lett.* **8** (2001), no. 1-2, 43-56.
- [5] C. Demeter, T. Tao and C. Thiele. Maximal multilinear operators, accepted for publication in *Trans. Amer. Math. Soc.*. Preprint available at arxiv.org/pdf/math.CA/0510581

- [6] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.* **31** (1977), 204-256.
- [7] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.* **11** (2001), no. 3, 465-588.
- [8] P. Halmos. Lectures in Ergodic Theory. Chelsea Publishing Co., New York, 1956.
- [9] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math.* (2) **161** (2005), no. 1, 397-488.
- [10] M. Lacey and C. Thiele. L^p estimates on the bilinear Hilbert transform for $2 < p < \infty$. *Ann. of Math.* (2) **146** (1997), no. 3, 693-724.
- [11] M. Lacey and C. Thiele. On Calderón's conjecture. *Ann. of Math.* (2) **149** (1999), no. 2, 475-496.
- [12] M. Lacey. The bilinear maximal functions map into L^p for $2/3 < p \leq 1$. *Ann. of Math.* (2) **151** (2000), no. 1, 35-57.
- [13] E. Lesigne. Sur la convergence ponctuelle de certaines moyennes ergodiques. *C. R. Acad. Sci. Paris Sér. I Math.* **298** (1984), no. 17, 425-428.
- [14] T. Ziegler, *Universal characteristic factors and Furstenberg averages*, J. Amer. Math. Soc. **20** (2007), no. 1, 53-97.

(C. Demeter) DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES CA 90095-1555

E-mail address: demeter@math.ucla.edu